47th NORTHEAST REGIONAL STOCK ASSESSMENT REVIEW COMMITTEE (SARC-47)

Report on the 2008 Summer Flounder (*Paralichthys dentatus*) Benchmark Stock Assessment

---

Mike Armstrong

Prepared for

Center for Independent Experts

The Centre for Fisheries and Aquaculture Science
Lowestoft Laboratory
Pakefield Road
Lowestoft
Suffolk NR33 0HT
England, United Kingdom

Phone: +44 1502 524362
Email: mike.armstrong@cefas.co.uk
www.cefas.co.uk

Contents

1. Background

This report provides an independent review of the assessment of summer flounder (*Paralichthys dentatus*) carried out at the Stock Assessment Workshops (SAW-47) and presented at the 47th Northeast Regional Stock Assessment Review Committee meeting. The Review Committee was provided with web access to stock assessment reports and background material prior to the meeting. I then participated in the 47th Northeast regional Stock Assessment Review Committee (SARC-47) meeting to review the assessments. This report includes my own review of the summer flounder assessment as well as required documentation including the Statement of Work, meeting Agenda and Terms of Reference.

2. Review activities

The Review Committee convened at the Laboratory of the Northeast Fisheries Science Center (NEFSC) in Woods Hole, Massachusetts, from June 16-20, 2008. The Committee comprised a chair and three panel members. Plenary sessions were open to the public.

A formal presentation of the Stock Assessment Workshop (SAW) results was given by the lead assessors from the Southern Demersal Working Group (SDWG), and specific issues were discussed. The assessors returned, when required, for further discussion and clarification of how the SAW Terms of Reference were addressed, including carrying out some additional model runs for clarification.

The panel members were then required to prepare an independent report indicating for each Term of Reference of the relevant SAW: i) whether the work that was presented is acceptable based on scientific criteria (e.g. consider whether the data were adequate and used properly, the analyses and models were carried out correctly, and whether the conclusions are correct/reasonable); and ii) whether the work provides a scientifically credible basis for developing fishery management advice.

The SARC chair and panel members prepared a first draft of the consensus report during the meeting. The panel members prepared their independent reports following the meeting. There were no disagreements between the panel members on any issues, and therefore my independent review given below to a large extent reflects the consensus report developed at the meeting, with additional comments. Some of the original Consensus Report text has been summarised, or expanded where appropriate, but without changing the Committee's agreed views.

3. Acknowledgements

I would like to thank all the SDWG members present at the meeting for their informative presentations of the SAW results and for providing helpful response to the SARC's questions. Many thanks also to staff at the Woods Hole Laboratory and particularly to Jim Weinberg and Paul Rago for their hospitality and help throughout the meeting. Many thanks also to the other members of SARC for productive discussions on the assessments.

4. Assessment of summer flounder

4.1 Overview

My overall conclusion for this stock is that the assessment and forecasts, as presented, provide a scientifically credible basis for developing fishery management advice for summer flounder. All Terms of Reference for this benchmark assessment were addressed by the SDWG.

The conclusion that spawning stock biomass and age composition have expanded to above any previous recorded values in the assessment time series, and that fishing mortality has declined to values lower than recorded in the time series, appears robust despite some issues with retrospective bias. The assessment is based on consistent results from a range of trawl surveys and an effective programme of commercial and recreational fishery sampling. Several forms of catch-at-age analysis (Adapt, ASAP and SS2) provided qualitatively similar results.

The main uncertainties in the assessment and advice include:

- The appropriate value assumed for natural mortality M. Expert opinion of the SDWG, backed by inferences from model fits, was for a higher M on males than females, and an overall higher combined-sex M than the value of 0.20 used previously. The revised value of M=0.25 (averaged over age classes) reduces the estimates of SSB_{MSY} and the ratio of SSB_{MSY} to current SSB. The Review Committee accepted the higher M value, but requested a number of additional runs at different M values so that the sensitivity to this assumption is fully transparent to managers.

- Retrospective bias in the assessment. This was pronounced during a recent period of rapid stock growth but appears reduced as the stock has stabilised. However the bias is indicative of a problem with historical data or assumed stock dynamics, and requires further investigation as the problem could occur again in the future.

- Possible biases in the assessment due to the use of a combined-sex assessment when there is sexual dimorphism, probable sex-related differences in mortality, and apparent changes in sex ratio over time.

I consider that the final assessment model represents a valid basis for developing management advice despite these uncertainties. However, future advice will benefit from further research on these issues.

The extent to which the assessment programme has addressed each of the Terms of Reference for the SAW is evaluated below.

4.2 ToR 1. Characterize the commercial and recreational catch, effort and CPUE, including descriptions of landings, discards and discard mortality

This Term of Reference was completed successfully. The data collection schemes are appropriate for estimating the quantity and size/age composition of all significant removals due

to commercial and recreational fishing. Sampling intensity has improved over time and summer flounder is considered one of the best sampled stocks off the NE coast. The fishery data can provide a suitable basis for exploring a range of catch-at-age models to provide credible fishery management advice.

The statistical catch-at-age models implemented using ASAP and SS2 for the assessment are able to provide a close fit between estimated and observed catches at age of summer flounder, and the fitted selection patterns are in accordance with past changes in technical measures affecting selectivity. Nonetheless, all the models show some retrospective bias, and there should be continued investigation of potential sources and magnitude of errors in the commercial and recreational data on retained and discarded fish as part of a broader evaluation of the causes of retrospective bias. This includes the assumed value of discard mortality, which will affect the assessment and stock status in the same way as the assumed value for natural mortality.

Sensitivity testing of the assessment to potential errors in different aspects of the fishery data could help target research effort where it will have greatest impact on improving the reliability of the assessment models. For example, the main issue with discard mortality may not be the absolute value but any trends over time due to changes in fishing practices.

Male and female summer flounder have demonstrably different rates of growth and maturity, and expert opinion is that natural mortality in males is higher than in females. There is evidence for changes in sex ratio in the population over time, which could cause bias in combined-sex assessment models and calculation of biological reference points. To address this properly would require sampling schemes to allow all fishery data and federal/state survey indices for summer flounder to be compiled separately by sex. This could provide better understanding of sources of bias in the assessments, or facilitate possible future exploration of sex-based assessments if warranted by the data.

It is recognised that collecting length and age composition data by sex is a potentially major undertaking for sampling the commercial and recreational fisheries. Provided the data are adequate, sex-disaggregated assessments are likely to be an improvement over sex-aggregated assessments where there are differences in biological parameters and population dynamics between sexes. However, there may be scope for using management strategy evaluation approaches using a range of operating models with sex-differentiated population dynamics, to investigate the potential risks of using combined-sex assessments and the cost-benefit of collecting fishery data by sex. The historical sex-disaggregated data from the NEFSC seasonal trawl surveys could facilitate such an exercise.

An important omission from the Southern Demersal Working Group reports to SARC-47 was an adequate characterization of the commercial and recreational fisheries. Information on the spatial distribution of fishing activities, changes in fishing effort over time, and changes in fishing gears and regulations affecting selectivity, would facilitate interpretation of the fishery data included in the assessment. Future SDWG reports should include a suitable Working Paper on fishery characterizations.

4.3 ToR 2. Review methods for using fishery-independent surveys as abundance indices in assessment models. *(a) Evaluate whether to combine several of the surveys into a composite survey index. If appropriate, implement this approach; (b) Develop and implement an appropriate statistical method to account for the probability of observing zeros in NEFSC survey tows*.

TOR 2(a) was addressed by the SDWG, and they made a valid conclusion (Working Paper 3) that the GLM analysis carried out by them is not a suitable basis for providing a combined tuning data set for the catch-at-age assessment. Such an analysis would only be statistically robust if the effect of location of individual tows was modeled (rather than just a survey effect), and if there were sufficient temporal and spatial overlap between the State and Federal surveys to allow the vessel/survey effects to be adequately determined. Unfortunately, the SDWG did not provide maps of survey distribution to show the extent of spatial overlap, but information available from other sources indicate that there may be insufficient overlap for this form of analysis. This means that any analysis would need to have an underlying model of seasonal and spatial dynamics of the population at each age to provide the necessary dynamic link between the surveys. However, if going down this rather complex route in the future, it would make more sense to include all the individual fishery and survey data sets in the framework of a spatially explicit assessment model rather than trying to develop some form of combined survey index external to the assessment model.

In the meantime there remains a potential problem in using any of the survey data sets unless there is evidence that the indices from each survey are directly proportional to population abundance at each age in the stock as a whole (i.e. constant catchability over time). For State surveys covering only part of the range of the stock, and particularly near the boundaries of the stock, it is possible that the abundance indices may not have a direct linear proportionality with overall population abundance. This could happen, for example if the distributional range of the stock expands and contracts with changes on overall abundance. Summer flounder has undergone substantial changes in abundance and age structure, and the assumption of constant catchability over time should not be assumed even if survey design remains constant. This applies equally to the larger scale NEFSC surveys. An expansion or contraction of geographic range of flounders in each age class into strata with different catchability could easily lead to changes in overall catchability. This could cause retrospective bias in the stock assessment during periods of rapid change in population size. Spatial variations in catchability could arise if there are between-stratum differences in trawl efficiency (e.g. due to depth-related changes in trawl-door spread or avoidance of the trawl). Raising factors for strata are also a potential source of error, if the strata have variable proportion of habitat suitable for trawling but have different densities of fish in the trawlable and non-trawlable areas. It is probably more productive for the SDWG to examine survey catchability issues than to look for simplistic ways to combine different survey data sets with different temporal and spatial coverage.

The wording of TOR 2(b) addresses occurrence of zero catches in tows, which is different than what the SDWG ultimately addressed. The SAW 47 TOR referring to ‘tows’ is different than the issue highlighted by the 2006 S&T assessment, which referred to zeros in the indices used in the assessment model. The comments below assume that the ToRs for the current assessment are wrongly worded. Nonetheless, if there are substantial numbers of zeros at the level of individual

tows, the SDWG might wish to consider survey analysis methods that model proportion of zero tows and catch rates at positive tows separately (e.g. delta-distribution approach). The relatively large number of zero indices for the older age classes at low stock size indicates that there must be a large proportion of zero tows for small non-zero indices, and the proportion is likely to have declined as the stock has expanded.

ToR 2(b), interpreted as referring to zero values for indices rather than tows, was addressed in detail by the SDWG. The evidence and analyses presented make a valid case for treating zeros in the survey indices as missing values. The SDWG thoroughly evaluated the suggestions and alternatives, such as using different small values, and showed that such approaches could lead to bias. Inserting small constant values is in effect adding wrong values that have no information on relative year class strength and may cause biases in the model parameter estimates. A more appropriate approach to examining the sensitivity of the assessment to missing values would be to fit a catch-curve type model to replace missing values with non-zero values more in line with the observed year class variations. Assessment model fits including these values, and then treating them as missing, would be more instructive than comparing missing values with small constants, no matter what "rule" is used for generating the constants. (Note: It is not suggested that replacements based on catch curves are used in the final assessment model.)

In general, there appears to be a number of different methods used for calculating indices of abundance from the Federal and State surveys (e.g. geometric or arithmetic mean catch rates). The SDWG should review the methods used for calculating abundance indices for the different State and Federal surveys with a view to ensure that appropriate and consistent methods are used, including treatment of tows with zero catches for individual age classes.

It is noted that many of the State surveys use the same age length keys (ALKs) derived from the larger-scale Federal surveys. The surveys are then treated as independent in the assessment, which is not strictly true due to some correlation of errors introduced by the common ALKs, particularly for larger fish where a length class may have several age classes. This could lead to the assessment appearing a bit more precise than it actually is. It may also lead to significant bias in the age composition of State survey catches if there are marked spatial variations in age composition of fish of a given length class.

4.4 ToR 3. Evaluate the feasibility of implementing alternative approaches to assess status of summer flounder stock and comment on any potential effects on estimates of F, SSB, and BRPs. Alternative approaches could consider:
a). Separate catch at age matrices for commercial and recreational fisheries, and resulting partial recruitment vectors for each fishery;
b). Regional differences (north, south) in catch at age matrices;
c). Potential gender differences in life span, growth rate, and natural mortality and implications of these factors for observed age- and length-specific sex ratios.
d). Strength of evidence for natural mortality rate used in the assessment; Update the estimate if appropriate.

Term of Reference 3 was completed and all of the suggested approaches in the ToR were addressed with varying degrees of success. The evaluation was sufficiently comprehensive to provide the basis for a final assessment model providing a credible basis for management advice.

The final ASAP model represents a logical methodological advance in allowing a statistical approach to modeling errors in the fishery and survey data. The configuration of the final two-fleet ASAP model provides a bridge to past ADAPT configurations, as it places considerable weight on the catch-at-age data. The results do not differ significantly from the equivalent single-fleet ADAPT runs. However the model provides a basis for future development that could further improve the management advice for the stock.

ToR 3(a): A number of multiple-fleet configurations of the ASAP and SS2 models were explored, with up to six separate catch-at-age matrices for commercial and recreational landings and discards. This has the potential for fitting time-varying selectivity patterns for different components of the fishery that have different selectivity characteristics. However, the exploration of multiple fleets suggests that the information content of historic sampling data may strongly constrain the number of catch-at-age matrices that can be modeled separately. The final assessment model chosen by SDWG included only two matrices, one for all commercial and recreational landings combined, and one for all discards irrespective of source. The suggested ToR option for two separate catch at age matrices, one for all commercial catches and one for all recreational fisheries, was not explored. The SDWG approach combines retained or discarded catch-at-age data from fishing methods that have very different characteristics and are subject to sampling schemes with different error characteristics and sampling rates (e.g. port sampling, commercial observer data, and MRFSS sampling of recreational catches). The SDWG should try to ensure that any combined catch-at-age data sets for which selectivity patterns are being estimated in ASAP or SS2 are as similar as possible in terms of fishery selectivity and error structure.

ToR 3(b): The SDWG carried out a useful descriptive exploration of spatial and temporal patterns in age compositions for the commercial landings and trawl surveys. Interpretation of the results is hindered by a lack of a detailed characterization of the fisheries. Any future developments of the assessment model to include spatial dynamics will require further evaluation of spatial patterns in data, spatial dynamics of the stock, and the accuracy of positional records in the fishery data.

ToR 3 (c) and (d): The SDWG provided a detailed evaluation of sex ratios, and differences in growth and maturity between males and female summer flounder, based on comprehensive data from NEFSC surveys. Unfortunately, data by sex are not available from the commercial and recreational fisheries, or from the State surveys which tend to cover areas closer inshore than the NEFSC offshore surveys. The SDWG evaluated the potential magnitude of natural mortality (M) in males and females based on published studies, comparisons with other fish stocks, and inferences from summer flounder data and model fits. The SDWG recommended an increase in the combined-sex natural mortality value from 0.20 to 0.25 on the basis of their evaluations. The Review Committee asked for some additional ADAPT and ASAP runs to investigate the implications of this for evaluation of current stock status in relation to BRPs, to ensure that the sensitivity to assumed M is fully transparent to managers.

The change in M resulted in a significant change in perception of stock status, and the decision to change M requires careful justification. The SDWG arguments for increased M included:

- M on males is likely to be higher than on females. This is based on observed maximum ages in male and female summer flounder, declines in proportion male with increasing age in surveys, inferences from exploratory sex-based SS2 model fits, and expert opinion of SDWG members familiar with studies on a wide range of stocks.
- Models with M values higher than 0.2 generally resulted in better fits to data.

The SDWG arrived at a weighted average M estimate of 0.25. This was calculated from values of M=0.3 in males and M=0.2 in females, inferred from recently observed maximum recorded ages of 14 years for females and 12 years for males in NEFSC surveys. A combined-sex M-schedule at age was developed by assuming these initial M rates by sex, an initial proportion of females at age 0 of 0.40 (from the NEFSC Fall survey), and population abundance decline over time at the sex-specific M rates. This gives slightly different M at each age, but with a mean of 0.25 for males and females combined.

The specific M by sex will be a function of the physiological determinants of longevity in males and females as well as the abundance of predators taking different sizes of summer flounder. The specific value for summer flounder at present cannot be determined from existing data whilst fishing remains a major determinant of average longevity, and because there are no data on sex ratio in commercial and recreational catches to determine if males have different fishing mortality at age than females due to the interaction of sexual dimorphism in growth and the selectivity of the fisheries. I do not have a basis for arguing against the SDWG's expert judgment in proposing a combined-sex value of M=0.25, provided that the sensitivity of stock status evaluations to choice of M is fully transparent to managers.

Comparative ADAPT and ASAP runs at M=0.2, 0.25 and 0.33 show that all runs lead to similar spawning stock biomass in 2007, but the B_{MSY} value derived from SSB per recruit curves and average recruitment declines substantially with increasing M. Hence, at M=0.25, the stock can be rebuilt by 2012 with higher $F_{rebuild}$ values and associated TALs than would be the case at M=0.20. However, the long-term maximum sustainable yield (at target F_{MSY} proxy of $F_{40\%}$) is lower at the higher M value, because a higher proportion of the fish production is lost to natural deaths. So there are trade-offs in the assumption of M.

Some aspects of the methods applied by SDWG to arrive at appropriate values of natural mortality require further consideration:

- Recent maximum ages in surveys will be strongly influenced by fishing, and also by differences in cumulative fishing mortality with increasing age in males and females;
- Sex ratios at the youngest ages could be biased in the NEFSC offshore surveys if the incidence of 0-gp and 1-gp fish within the survey area is dependent on migration offshore from inshore nursery areas, and if the proportion moving offshore depends on biological processes such as maturation which varies between sexes (this could also lead to biased estimates of proportion mature in young flounders). The youngest age group in the NEFSC surveys is substantially under-represented compared with older age classes, and the indices correlate very poorly with those of the older ages in each cohort. This implies that the NEFSC survey data for the youngest age groups may not be representative of the population at these ages, and any inferences on parameters such as sex ratio, proportion mature and mean length at age should be considered as potentially biased by distributional effects. A more comprehensive spatial and seasonal analysis of sex ratio at age should be carried out at the scale of individual survey tows, to evaluate potential biases in estimates from surveys. Sex ratio and maturity data should be collected in the inshore State surveys to further investigate this issue. During autumn, spawning aggregations with skewed sex ratios may be prevalent.

4.5 T0R 4. Compare results from alternative modeling approaches with those from the VPA model, to evaluate the robustness of VPA model results. Perform retrospective analyses of F, SSB, and recruitment for the models, and describe potential effects of retrospective patterns on assessment and rebuilding.

Comparison of results from different modelling approaches

This ToR was in general adequately addressed by the SDWG. The ToR required alternative models to the ADAPT VPA to be run for comparison as a means of evaluating the robustness of the VPA results. All the models evaluated by SDWG (ADAPT, ASAP, SS2) gave similar stock trends when configured similarly. The final choice of a statistical catch-at-age model (implemented using the ASAP package) provides a suitable bridge with the previously adopted ADAPT model, whilst providing greater flexibility for accommodating the types of errors inherent in the input catch and survey data. The ability of ASAP (and SS2) to model selectivity in a coherent manner by fleet is an advantage over ADAPT for which selectivity is whatever can be inferred from the variable F-at-age values obtained treating the catch-at-age data as exact. ASAP and SS2 also have the advantage in principle of being able to produce more consistent integration of assessments, reference point calculation and projections and, importantly, better defined estimates of uncertainty.

Although the sequence of model-fitting steps for each type of model was clearly presented in the SDWG assessment report, it was difficult to properly review the appropriateness of decisions made by the SDWG on the basis of model diagnostics. During the meeting, therefore, time was spent with the SDWG working through the diagnostics available to the SDWG during the model development process (many thanks are due in particular to Mark Terceiro for help in this respect). The process and decisions made by SDWG appeared appropriate. However, future

assessment reports should contain a more transparent presentation of model specification and model building/selection procedures, including relevant diagnostics supporting decisions made, to facilitate review and enhance credibility. Suitable material could include:

- A succinct description of the process, sensitivity runs, etc, given in the main assessment report, including a reference table summarizing the differences in model structure, settings and input data for each run.
- A separate working document providing more detail of model development and selection. This should include:

- A clear graphical presentation of diagnostics such as plots of residuals for surveys and catch-at-age data by year, and of estimates and errors of parameters, ogives, etc. (present only the diagnostics necessary to explain model selection decisions).
- A tabulated summary of model fits showing likelihood components and information relevant to interpretation of the comparison.
- Where MCMC is used, a summary of information on priors used, posterior distributions and MCMC performance.

It may be useful to consider development of standardized methods for display of outputs (e.g., using common R or similar scripts).

Ultimately, the choice of assumed M had greater influence on management-related variables than the type of model used or the decisions on alternative fleet structure, error structures, etc. in the different models This conclusion was arrived at after the Review Committee requested additional model runs to examine the dependency of management-related outputs to a) data updating, b) model assumptions and changes within the ADAPT update, and c) use of the recommended integrated catch-at-age model implemented using the ASAP package (see ToR 7 discussion)..

Retrospective analyses

Previous assessments of summer flounder have displayed strong retrospective patterns while the stock has been expanding rapidly. The different models explored by SDWG all still display similar patterns, although arguably less severe in the adopted ASAP base case. All models show a reduction in retrospective bias for the past three years. The NEFSC surveys indicate a larger relative increase in biomass in the 2000s than given by the converged ADAPT population estimates or the estimates from ASAP and SS2. Diagnostics requested by the Review Committee showed very large positive residuals between observed and fitted survey indices for several age classes around the time that biomass was increasing towards the recent maximum in the NEFSC surveys. The residual errors are more pronounced at some intermediate ages. Residual errors have since decreased, as has retrospective bias, as the stock has apparently stabilized.

Potential reasons for these patterns include:

- Departures from the assumption of direct proportionality between survey indices and true abundance when abundance and spatial extent of the stock is rapidly expanding;

- Trends in unaccounted-for removals from the stock (e.g. trends in any biases in commercial or recreational catch estimates, or trends in discard mortality or natural mortality).

The different assessment model runs requested by the Review Committee to examine sensitivity to assumed natural mortality, gave roughly the same population abundance in 2007 (in absolute values), but higher historical values, when larger values of M were used. Progressively higher M values gave a more "domed" biomass trend in the 2000s compared to the more plateau shaped trend given at lower M values. This may explain why higher M values now appear to give better model fits in ASAP (based on likelihood profile for different M values provided during the review meeting). This warrants further investigation through retrospective analysis. In this context, changes in assumed values of M and discard mortality could have analogous effects.

The underlying cause of retrospective patterns should ideally be determined and then accounted for directly in the assessment, so that the characterization of uncertainty in parameters and derived values is integrated in the model. It is not appropriate (see ToR 8) to attempt to "fix" projections to account for retrospective pattern; and the SDWG decision not to do this is appropriate.

4.6 ToR 5. Based on the "best" model or models, estimate fishing mortality rate, recruitment, spawning stock biomass, and total stock biomass for the current year and characterize the uncertainty of those estimates. If possible, also include estimates for earlier years with uncertainty estimates.

ToR 5 was successfully addressed. The SDWG adopted a base-case assessment model using ASAP and provided estimates of F, R and SSB in 2007 (the terminal year). Time series of these quantities were also produced. The SDWG also attempted to characterize uncertainty in the estimates of F and SSB from the base case assessment for use in projection (ToR 8) and, differently, for earlier years. However, the method adopted is unlikely to fully characterize the uncertainty and does not maintain coherency between all components of the assessment and forecast procedure. This approach to integrated assessment and projections is possible within ASAP but is understood not to have been used due to management requirements for outputs produced using AGEPRO software.

The derived estimates and calculated uncertainties appear appropriate as a basis for determining stock status, making projections and providing management advice. However, better within-model uncertainty estimates should be determined. An evaluation of the sensitivity of the assessment to major sources of uncertainty (especially M) should also have been provided.

4.7 ToR 6. Examine and evaluate the role of the environment on past and present summer flounder recruitment success.

This ToR was adequately addressed. Evidence presented in the two working papers (WP11, WP12) suggests that both temperature and the North Atlantic Oscillation (NAO) may correlate

with summer flounder recruitment. However I agree with the decision of the SDWG not to consider including environmental factors in current model configurations.

This should not preclude continued study of how the productivity of this (and other stocks) may be affected by random and persistent changes in environmental conditions, and carrying out hypothesis-driven field studies to determine the underlying biological processes.

4.8 ToR 7 Biological Reference Points: a) Update or redefine biological reference points (BRPs; proxies for B_{MSY} and F_{MSY}), taking into account conclusions from earlier assessments and findings from TOR 6 (i.e., recruitment and the environment). Estimate uncertainty in BRPs. Comment on the scientific adequacy of existing and redefined BRPs; b) Evaluate current stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 7a).

ToR 7a): The SDWG successfully addressed ToR 7a in terms of providing re-defined estimates of biological reference points for the base case assessment. I agree with the SDWG decision to move away from F_{max} as a proxy for F_{MSY} target reference point due to poor definition of the F_{max} value. (Furthermore, the revised yield-per-recruit curve indicates relatively small marginal increases in yield per recruit for quite large increases in F approaching F_{max} .) However, the requirement to provide measures of uncertainty in the reference points was not addressed by the SDWG. The Review Committee required more clarity on how the adoption of the new base-case ASAP model, rather than a simple update of the 2006 ADAPT VPA using data from 2006-2007, affected management-related outputs including the perception of stock status relative to biological reference points. The SDWG therefore undertook additional assessments at the Panels' request.

These runs are described in some detail in the Consensus report, but are repeated below as there is a CIE requirement for individual independent reviews to stand alone and not refer readers to the Consensus Report for essential details. The results of the different runs are tabulated in **Appendix 1**.

The first three columns of the table in Appendix 1 show MSY-related reference points and quantities of management interest for an assessment using the 2006 ADAPT VPA updates (S&T 2006 series), but using three different values for M (M=0.20, as used previously in ADAPT; M=0.25 as used in this year's final ASAP assessment, and M=0.33 to help illustrate the sensitivity to choice of M).

The next three columns show the same quantities for an ADAPT run similar to the 2006 model but using inputs updated to include 2006-2007 data (T 2007 series).

The third set of three columns shows the same quantities for the SDWG-adopted base case ASAP assessment using M=0.25 (middle column, labeled **F08_T2007_T2**) and sensitivity runs using M of 0.2 and 0.33.

The Consensus Report gives a very detailed evaluation of these runs, but the essential results can be summarised as follows:

- The results indicate that adoption of the ASAP base case instead of ADAPT is not the cause of the revision in reference points proposed by SDWG, or the revised estimates of

F, SSB and stock status. (Comparing the ADAPT 2007 T series and the ASAP results, the estimates of reference points and quantities of management interest are almost the same for any given M assumption.)

- The inclusion of data for 2006 and 2007, the changes in the weight-at-age and treatment of partial recruitment, and crucially the adoption of a larger value for M, are the main causes of the revised perception of stock status.
- F_{max} is not a robust proxy for F_{MSY} due to a progressively more asymptotic yield-per-recruit curve as M is increased. Proxies for thresholds and targets such as $F_{35\%}$ and $F_{40\%}$ appear more robust.
- Increasing the value of assumed M results in decreased estimates of SSB_{MSY} but relatively little change in estimated SSB_{term} (SSB in terminal year), i.e. current stock size is still below, but closer to the target value for rebuild. At M=0.20, the point estimate of F in 2007 is above both $F_{35\%}$ and $F_{40\%}$. At M=0.25 and 0.33, F in 2007 is below the $F_{35\%}$ threshold but above the $F_{40\%}$ target.

At a time when the stock is close to rebuilding targets, and perceptions of status are critically dependent on detailed assessment choices and assumptions, it is important to provide for as much stability as possible in management and assessment processes so as not to confuse data-driven signals with changed analytical choices. I therefore agree with the SDWG's use of selectivity ogives fitted in a statistical catch-at-age model (ASAP), to provide a more stable basis for the "non-parametric" reference point calculation.

However, using a short running average of weight at age can introduce instability to reference point estimation. It is recommended that future updates to biological reference points should consider methods to stabilize weight-at-age values for use in reference point calculation. (NB this recommendation applies to reference point calculation, not necessarily to weights at age used for short term forecasts or current SSB calculation).

ToR 7b: The evaluation of stock status using reference points derived from the base case ASAP assessment is an appropriate basis for developing management advice for summer flounder.

The "existing biological reference points" (as agreed in 2006) based on the previous ADAPT formulation are not valid for comparison with the base case ASAP assessment due to the changes in assessment model and yield per recruit inputs this year (particularly the revised value of M).

Overall, as stated by the SDWG based on the base case assessment, the stock is seen to be rebuilding towards the target SSB value, and F has been reduced to close to the target value. The stock is not overfished and overfishing is not occurring, but the stock is still rebuilding.

It should be remembered that the reference points and terminal F and SSB values are in fact subject to error. The SDWG did not provide probability profiles of F/F_{MSY} or SSB/SSB_{MSY} for any assessments. Such information would be useful and can be derived from the MCMC procedure.

4.9 ToR 8. Stock Projections:.a) Recommend what modeling approaches and data should be used for conducting single and multi-year stock projections, computing TACs or TALs, and measures of uncertainty; b) If possible: i. Provide numerical examples of short term projections (2-3 years) of biomass and fishing mortality rate, and characterize their uncertainty, under various TAC/F strategies and ii. Compare projected stock status to existing rebuilding or recovery schedules, as appropriate.

Although all aspects of **TORs 8 (a) and (b)** were addressed by SDWG, there were a number of technical issues with the procedures used for characterising the uncertainty of the forecasts. Clarification on the procedures adopted was sought from SDWG members during the review meeting, and it is considered likely that the percentiles around the forecasted values are underestimates of the true uncertainty. This does not invalidate the management advice based on point values in the forecast. However it means that the practice of giving a 25^{th} percentile on TAL in the forecast to provide some buffer against retrospective bias in the assessment is based on a figure which may not accurately reflect the "true" 25^{th} percentile. (Without carrying out a full management strategy evaluation to evaluate risk associated with different harvest control rules based on error-prone assessment results, the choice of any percentile as a "buffer" is somewhat arbitrary anyway.)

Although the AGEPRO program generates stochastic variation in terminal year abundance (measurement error) and subsequent recruitment (including process error), the coherent relationship among parameters is lost using this approach. The SDWG should base stock projections on the MCMC output of projection parameters/variables to ensure appropriate linkage between the uncertainties in the assessment and those in the forecast.

ToR 8(b) i: The SDWG provided the minimal sufficient projections under the premise that additional projections evaluating alternative scenarios would be prepared at a later date in response to recommendations by the SSC and the Council. Essentially, therefore, it was not possible for the SARC to fully address this TOR. SARC can only evaluate the small number of runs and 1-year projected figures as provided

ToR 8(b) ii: The Review Committee noted that a specific annual rebuilding schedule was not clearly presented. It is apparent that the guiding factor at this point is the $F_{rebuild}$ that will allow rebuilding to SSB_{MSY}, defined based on the corresponding biological reference points, by the end of 2012. Based on the figures presented, the stock could rebuild at both the proposed $F_{rebuild}$ and proposed F_{target} by the end of 2012.

For $F_{rebuild}$ and F_{target} strategies, short term projections of SSB and potential yield (TAL) are provided for one year in the future (2009). No catch projections for further years were provided. Future assessments should include a fuller set of constant catch and constant exploitation options and report the probability of rebuilding within the required timeframe and expected year to achieve rebuilt status for each option.

5 Research recommendations

5.1 ToR 9. Review, evaluate and report on the status of the Research Recommendations offered in recent SARC reviewed assessments and in the 2006 "Methot" Review.

The SDWG did an excellent job of addressing research recommendations from the previous SARC review and the 2006 Methot review. The SARC 47 review was greatly assisted by having all research recommendations in a single section.

Research prioritization could be improved through appropriate simulation and sensitivity analyses constructed to identify factors that significantly affect assessment outputs relevant to management advice. Those factors that most affect advice should then be classified as high priorities. For example, the sensitivities presented in this assessment suggest that natural mortality, including differences in natural mortality between males and females, is an important issue that significantly affects assessment conclusions and management advice.

5.2. Summary of SARC review recommendations

Recommendations given in this report (relevant sections of report are indicated)

1. Future SDWG reports should include a suitable Working Paper on fishery characterizations.(4.2)

2. SDWG might wish to consider survey analysis methods that model proportion of zero tows and catch rates at positive tows separately (e.g. delta-distribution approach). (4.3)

3. The SDWG should review the methods used for calculating abundance indices for the different State and Federal surveys with a view to ensure that appropriate and consistent methods are used, including treatment of tows with zero catches for individual age classes. (4.3)

4. The SDWG should try to ensure that any combined catch-at-age data sets for which selectivity patterns are being estimated in ASAP or SS2 are as similar as possible in terms of fishery selectivity and error structure. (4.4)

5. A more comprehensive spatial and seasonal analysis of sex ratio at age should be carried out at the scale of individual survey tows, to evaluate potential biases in estimates from surveys. Sex ratio and maturity data should be collected in the inshore State surveys to further investigate this issue. (4.4)

6. Future assessment reports should contain a more transparent presentation of model specification and model building/selection procedures including relevant diagnostics supporting decisions made, to facilitate review and enhance credibility. Consider development of standardized methods for display of outputs (e.g. using common R or similar scripts. (4.5)

7. Future updates to biological reference points should consider methods to stabilize weight-at-age values for use in reference point calculation. (4.8)

8. Stock projections should be based on the MCMC output of projection parameters/variables to ensure appropriate linkage between the uncertainties in the assessment and those in the forecast. (4.9)

Review Committee recommendations

The Review Committee's suggestions for high priority research recommendations for summer flounder are given below. I support these, but have added some observations in the text below the recommendations.

1. Continued evaluation of natural mortality and the differences between males and females. This should include efforts to estimate natural mortality, such as through mark-recapture programs and telemetry.
2. Continue efforts to improve understanding of sexually dimorphic mortality and growth patterns. This should include monitoring sex ratios and associated biological information in the fisheries and all ongoing surveys to allow development of sex-structured models in the future.
3. Conduct sensitivity analyses to identify potential causes of the recent retrospective pattern. Efforts should focus on identifying factors in both survey and catch data that could contribute to the decrease in cohort abundance between initial estimates based largely on survey observations and subsequent estimates influenced by fishery dependent data as the cohort recruits to the fishery.
4. Develop methods that more fully characterize uncertainty and ensure coherence between assessments, reference point calculation and projections.

In relation to Panel recommendation (2), it is further emphasized that a much clearer understanding is needed of how the perception of sex ratio (or proportion mature at age) given by surveys could be affected by differences in catchability of the different population components included in these ratios. If the youngest fish are more concentrated close inshore, but subsequently exhibit changes in depth distribution associated with their size or with the onset of maturity, this could easily lead to very biased perceptions of sex ratio and proportion mature in young summer flounder caught in the offshore NEFSC surveys. The different behaviour patterns of males and females during the spawning season can also lead to skewed sex ratios at small spatial scales, so time-of-year should be a factor in any study, and data should be viewed at the scale of individual tows.

In relation to Review Committee recommendation (3), I would like to emphasizes that although the magnitude of retrospective bias appears to have declined in recent years as stock abundance has stabilised, the bias could increase again in the future. The evaluation of retrospective bias should cover all aspects of data and assumed dynamics that could reasonably be considered as potentially significant sources of bias. This includes the assumption of constant catchability in surveys, possible biases in commercial and recreational catch estimates, and potential for trends

in population parameters assumed to be constant over time (e.g. natural mortality and discard mortality).

6. Alternative Biological Reference Points

The SARC 47 Review Panel accepted the alternative biological reference points as recommended by the SDWG and given in the assessment report.

7. Independent analyses conducted by the CIE reviewer.

No independent analyses were carried out by me other than some basic exploration of the internal consistency of the survey data sets which it is unnecessary to repeat here.

8. Additional questions not in the Terms of Reference.

No additional questions were raised during the SARC meeting.

APPENDIX 1. Results of a range of different assessment model runs requested by the review Panel

	ADAPT VPA S&T 2006	ADAPT VPA S&T 2006	ADAPT VPA S&T 2006	ADAPT VPA T2007	ADAPT VPA T2007_M25	ADAPT VPA T2007_M33	ASAP F08_T2007_T2_M20	ASAP F08_T2007_T2	ASAP F08_T2007_T2_M33
NON-PARAMETRIC	(deterministic)	(deterministic)	(deterministic)	(stochastic)	(stochastic)	(stochastic)	(stochastic)	(stochastic)	(stochastic)
	M = 0.20	mean M=0.25	mean M=0.33	mean M=0.20	mean M=0.25	mean M=0.33	mean M=0.20	mean M=0.25	mean M=0.33
Fmax	0.280	0.372	0.462	0.419	0.604	1.769	0.393	0.558	1.710
MSY (mt)	21,444	19,096	17,372	14,629	13,120	10,155	16,834	12,868	10,967
SSBmax(mt)	89,411	65,606	53,650	53,384	39,314	18,489	61,653	38,547	20,973
Fterm	0.410	0.520	0.527	0.311	0.311	0.317	0.300	0.288	0.290
Yterm	13,779	13,779	13,779	10,368	10,368	10,368	10,368	10,368	10,368
SSBterm	47,498	41,449	42,441	42,142	42,919	43,711	42,185	43,363	44,066
Fterm/Fmax	1.46	1.40	1.14	0.74	0.51	0.18	0.76	0.52	0.17
Yterm/MSY	0.64	0.72	0.79	0.71	0.79	1.02	0.62	0.81	0.95
SSBterm/SSBmax	0.53	0.63	0.79	0.79	1.09	2.36	0.68	1.12	2.10
F35%	0.218	0.265	0.291	0.281	0.337	0.379	0.263	**0.310**	0.352
MSY (mt)	21,429	18,715	16,934	14,767	13,389	12,055	16,974	13,122	12,026
SSB35%(mt)	109,994	85,127	74,639	73,624	60,333	54,061	85,570	**60,074**	53,811
Fterm	0.410	0.520	0.527	0.311	0.311	0.317	0.300	0.288	0.290
Yterm	13,779	13,779	13,779	10,368	10,368	10,368	10,368	10,368	10,368
SSBterm	47,498	41,449	42,441	42,142	42,919	43,711	42,185	43,363	44,066
Fterm/F35%	1.88	1.96	1.81	1.11	0.92	0.84	1.14	0.93	0.82
Yterm/MSY	0.64	0.74	0.81	0.70	0.77	0.86	0.61	0.79	0.86
SSBterm/SSB35%	0.43	0.49	0.57	0.57	0.71	0.81	0.49	0.72	0.82
F40%	0.183	0.220	0.238	0.234	0.276	0.307	0.219	**0.255**	0.285
MSY (mt)	20,837	18,163	16,385	14,480	13,070	11,551	16,632	12,807	11,515
SSB40%(mt)	125,723	97,306	85,325	84,306	69,133	60,907	98,024	68,743	60,016
Fterm	0.410	0.520	0.527	0.311	0.311	0.317	0.300	0.288	0.290
Yterm	13,779	13,779	13,779	10,368	10,368	10,368	10,368	10,368	10,368
SSBterm	47,498	41449	42441	42,142	42,919	43,711	42,185	43,363	44,066
Fterm/F40%	2.24	2.36	2.21	1.33	1.13	1.03	1.37	1.13	1.02
Yterm/MSY	0.66	0.76	0.84	0.72	0.79	0.90	0.62	0.81	0.90
SSBterm/SSB40%	0.38	0.43	0.50	0.50	0.62	0.72	0.43	0.63	0.73

Appendix 2: Review Committee members

John Carmichael, chair
Mike Armstrong
Kevin Stokes
Yan Jiao

Appendix 3: Terms of reference (from Annex 1, Statement of Work)

No changes to these were made prior to or during the SARC 47 review meeting.

DRAFT Assessment Terms of Reference for SAW/SARC-47 in June, 2008

(Last Revised: Sept. 27, 2007)

Summer flounder

1. Characterize the commercial and recreational catch, effort and CPUE, including descriptions of landings, discards and discard mortality.

2. Review methods for using fishery-independent surveys as abundance indices in assessment models.

   a. Evaluate whether to combine several of the surveys into a composite survey index. If appropriate, implement this approach.

   b. Develop and implement an appropriate statistical method to account for the probability of observing zeros in NEFSC survey tows.

3. Evaluate the feasibility of implementing alternative approaches to assess status of summer flounder stock and comment on any potential effects on estimates of F, SSB, and BRPs. Alternative approaches could consider:

   a. Separate Catch at age matrices for commercial and recreational fisheries, and resulting partial recruitment vectors for each fishery.

   b. Regional differences (north, south) in catch at age matrices.

   c. Potential gender differences in life span, growth rate, and natural mortality and implications of these factors for observed age- and length-specific sex ratios.

   d. Strength of evidence for natural mortality rate used in the assessment; Update the estimate if appropriate.

4. Compare results from alternative modeling approaches with those from the VPA

model, to evaluate the robustness of VPA model results. Perform retrospective analyses of F, SSB, and recruitment for the models, and describe potential effects of retrospective patterns on assessment and rebuilding.

5. Based on the "best" model or models, estimate fishing mortality rate, recruitment, spawning stock biomass, and total stock biomass for the current year and characterize the uncertainty of those estimates. If possible, also include estimates for earlier years with uncertainty estimates.

6. Examine and evaluate the role of the environment on past and present summer flounder recruitment success.

7. Biological Reference Points

   a. Update or redefine biological reference points (BRPs; proxies for B_{MSY} and F_{MSY}), taking into account conclusions from earlier assessments and findings from TOR 6 (i.e., recruitment and the environment). Estimate uncertainty in BRPs. Comment on the scientific adequacy of existing and redefined BRPs.

   b. Evaluate current stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 7a).

8. Stock Projections

   a. Recommend what modeling approaches and data should be used for conducting single and multi-year stock projections, computing TACs or TALs, and measures of uncertainty.

   b. If possible,

      i. Provide numerical examples of short term projections (2-3 years) of biomass and fishing mortality rate, and characterize their uncertainty, under various TAC/F strategies and

      ii. Compare projected stock status to existing rebuilding or recovery schedules, as appropriate.

9. Review, evaluate and report on the status of the Research Recommendations offered in recent SARC reviewed assessments and in the 2006 "Methot" Review.

Appendix 4: Draft Agenda 47th Northeast Regional Stock Assessment Workshop (SAW 47). Stock Assessment Review Committee (SARC) Meeting

Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

June 16 - 20, 2008 Sessions are open to the public, except where indicated.

TOPIC	PRESENTERS	RAPPORTEURS

Monday, 16 June (1:00 – 5:00 PM)……..……

Welcome	**James Weinberg**, SAW Chairman
Introduction	**John Carmichael**, SARC Chairman
Agenda	
Conduct of Meeting	
Summer flounder (A)	**M. Terceiro**, **J. Coackley, M. Maunder**
	Rich Wong
SARC Discussion	**John Carmichael**

Tuesday, 17 June (9 AM – Noon)……..……

Summer flounder (A) – finish presentations.	
	M. Terceiro, **J. Coackley, M. Maunder**
	Rich Wong
SARC Discussion	**John Carmichael**

Tuesday, 17 June (1:15 PM – 5 PM)……..……

Q&A #1 between Reviewers and All Presenters, clarification of any issues. (Open Meeting)	
	Rich Wong
SARC Discussion	**John Carmichael**

Wednesday, 18 June (9 AM – Noon) ……..…..……
SARC Panel deliberations/report writing (Closed Meeting).

Wednesday, 18 June (1:15 PM – 4 PM)……..…..……

Q&A #2 between Reviewers and All Presenters, clarification of any issues. (Open Meeting)	
	Rich Wong
SARC Discussion	**John Carmichael**

Wednesday, 18 June (4 PM – 5 PM) ……..…..……
SARC Report writing (Closed Meeting).

Thursday, 19 June (and possibly 20 June. AM)……..……
SARC Report writing (Closed Meeting).

Appendix 5: Bibliography

Working Papers Prepared in Support of SARC 47 Terms of Reference

#	Title	Author	
1	Estimation of Commercial Fishery Discards of Summer Flounder: Update 2007 or Revise the 1989-2007 Time Series?	anon.	
2	Discard Mortality of Summer Flounder in the Inshore Trawl Fishery	Emerson Hasbrouck Tara Froehlich Kristin Gerbino John Scotti	
3	Some Approaches to the Integration of Survey Abundance Indices used in VPA Calibration	Mark Terceiro	
4	Simulation Studies of Issues Associated with Filling Zeros in VPA Tuning Indices	Chris Legault Al Seaver	
5	Some More Thoughts on Filling Zeros in Tuning Indices: A Simple Regression Example	Chris Legault	
6	The Treatment of "Zero" Observations in the Summer Flounder ADAPT VPA Calibration	Mark Terceiro	
7	Evaluation of summer flounder life history parameters from NEFSC trawl survey data, 1992 – 2006.	Jeffrey C. Brust	
8	A Review of Natural Mortality of Summer Flounder	Rich Wong	
9	Analysis of Trends in Sex Ratio, Implications for Natural Mortality, and Variation in Age-Length Keys in Summer Flounder	Eric N. Powell Jason Morson	
10	Re-evaluation of Summer Flounder (*Paralichthys dentatus*) Stock Status Following Adjustments for Retrospective Bias and Inclusion of Trophic Effects	Victor Crecco	
11	Modeling environmental factors and summer flounder recruitment success	Mark Terceiro	
12	Wavelet Analysis of Trends in Summer Flounder YOY and Spawner-Recruit Relationships	Eric Powell	
13	Specifying Initial Conditions for Forecasting When Retrospective Pattern Present	Chris Legault and Mark Terceiro	

Appendix 6: Statement of Work

Attachment A: Statement of Work for Dr. Michael Armstrong

External Independent Peer Review by the Center for Independent Experts

SARC 47: Summer Flounder Benchmark Stock Assessment

Meeting Date: June 16 – 20, 2008

Statement of Work (SOW) for CIE Panelists
(including a description of SARC Chairman's duties)

General

The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development (SAW Working Groups or ASMFC technical committees), assessment peer review, public presentations, and document publication.

The SARC47 review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the South Atlantic Fishery Management Council. The panel will convene at the Woods Hole Laboratory of the Northeast Fisheries Science Center (NEFSC) in Woods Hole, Massachusetts during June 16-20, 2007 to review one assessment (Summer flounder, *Paralichthys dentatus*). In the days following the review of the assessment, the panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Overview of CIE Peer Review Process

The Office of Science and Technology implements measures to strengthen the National Marine Fisheries Service's (NMFS) Science Quality Assurance Program (SQAP) to ensure the best available high quality science for fisheries management. For this reason, the NMFS Office of Science and Technology coordinates and manages a contract for obtaining external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of stock assessments and various scientific research projects. The primary objective of the CIE peer review is to provide an impartial review, evaluation, and recommendations in accordance to the Statement of Work (SoW), including the Terms of Reference (ToR) herein, to ensure the best available science is utilized for the National Marine Fisheries Service management decisions.

The NMFS Office of Science and Technology serves as the liaison with the NMFS Project Contact to establish the SoW which includes the expertise requirements, ToR, statement of tasks for the CIE reviewers, and description of deliverable milestones with

dates. The CIE, comprised of a Coordination Team and Steering Committee, reviews the SoW to ensure it meets the CIE standards and selects the most qualified CIE reviewers according to the expertise requirements in the SoW. The CIE selection process also requires that CIE reviewers can conduct an impartial and unbiased peer review without the influence from government managers, the fishing industry, or any other interest group resulting in conflict of interest concerns. Each CIE reviewer is required by the CIE selection process to complete a Lack of Conflict of Interest Statement ensuring no advocacy or funding concerns exist that may adversely affect the perception of impartiality of the CIE peer review. The CIE reviewers conduct the peer review, often participating as a member in a panel review or as a desk review, in accordance with the ToR producing a CIE independent peer review report as a deliverable. The Office of Science and Technology serves as the COTR for the CIE contract with the responsibilities to review and approve the deliverables for compliance with the SoW and ToR. When the deliverables are approved by the COTR, the Office of Science and Technology has the responsibility for the distribution of the CIE reports to the Project Contact.

Requirements for CIE Reviewers

CIE reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models and Biological Reference Points. Expertise should include both the use of statistical catch-at-age and traditional VPA approaches. Experience with comparative studies of these approaches is especially valuable. Reviewers should also have experience in evaluating measures of model fit, identifiability, uncertainty, and forecasting. Experience with flatfish population dynamics would be useful.

Specific Activities and Responsibilities

The CIE's deliverables shall be provided according to the schedule of milestones listed on Page 6**.** The CIE reviewers, along with input and leadership from the SARC Chairman, will write the SARC Summary Report. In addition, each CIE reviewer will write an individual independent review report. These reports will provide peer-review information for a presentation to be made by NOAA Fisheries at meetings of the New England and Mid-Atlantic Fishery Management Councils in 2008. The SARC Summary Report shall be an accurate representation of the SARC panel viewpoint on how well each SAW Term of Reference was completed (please refer to Annex 1 for the SAW Terms of Reference).

The three CIE reviewers shall conduct an impartial and independent peer review in accordance with the Terms of Reference (ToR) herein. The three SARC CIE reviewers' duties shall occupy a maximum of 14 days per person (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; and several days following the open meeting to contribute to the SARC Summary Report and to produce the Independent CIE Reports).

Not covered by the CIE, the SARC chair's duties shall occupy a maximum of 15 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation.)

Charge to SARC panel

The panel is to determine and write down whether each Term of Reference of the SAW (see Annex 1) was or was not completed successfully during the SARC meeting. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. Where possible, the chair shall identify or facilitate agreement among the reviewers for each Term of Reference of the SAW.

If the panel rejects any of the current Biological Reference Point (BRP) proxies for B_{MSY} and F_{MSY}, the panel should explain why those particular proxies are not suitable and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs are the best available at this time.

Roles and responsibilities

(1) Prior to the meeting

(SARC chair and CIE reviewers)
Review the reports produced by the Working Groups and read background reports.

(2) During the Open meeting

(SARC chair)
Act as chairperson, where duties include control of the meeting, coordination of presentations and discussion, making sure all Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For the assessment, review both the Assessment Report and the Assessment Summary Report.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)
For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. From a reviewer's point of view, determine whether each Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(3) After the Open meeting

(SARC CIE reviewers)
Each reviewer shall prepare an Independent CIE Report (see Annex 2). This report should explain whether each Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified above in the "Charge to SARC panel" statement.

If any existing Biological Reference Point (BRP) proxies are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report produced by each reviewer.

The Independent CIE Report can also be used to provide greater detail than the SARC Summary Report on specific Terms of Reference or on additional questions raised during the meeting.

(SARC chair)
The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was adequate to complete the Terms of Reference of the SAW. If

appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report.

(SARC chair and CIE reviewers)
The SARC Chair and CIE reviewers will prepare the SARC Summary Report. Each CIE reviewer and the chair will discuss whether they hold similar views on each Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar or a consensual view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion.

The SARC Summary Report (please see Annex 3 for information on contents) should address whether each Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

If any existing Biological Reference Point (BRP) proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

Schedule

The milestones and schedule are summarized in the table below. No later than July 7, 2008, the CIE reviewers shall submit their Independent CIE Reports to the CIE lead coordinator Mr. Manoj Shivlani via e-mail to shivlanim@bellsouth.net

and CIE regional coordinator Dr. David Sampson via e-mail to David.Sampson@oregonstate.edu.

Milestone	Date
CIE reviewers attend the SARC workshop to conduct peer review at Northeast Fisheries Science Center (NEFSC) in Woods Hole, MA, USA	June 16-19
SARC Chair and CIE reviewers work at the NEFSC drafting reports	June 19-20
Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair **	July 7
CIE reviewers submit Independent CIE Reports to CIE for approval	July 7
SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)	July 14
CIE provides reviewed Independent CIE Reports to NMFS COTR for approval	July 21
COTR notifies CIE of approval of reviewed Independent CIE Reports	July 28, 2008 *
COTR provides final Independent CIE Reports to NEFSC contact	July 28, 2008

* Assuming no revisions are required of the reports.
** The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

NEFSC Contact person and SAW Chairman:
Dr. James R. Weinberg, NEFSC, Woods Hole, MA. 508-495-2352,
James.Weinberg@noaa.gov

Submission and Acceptance of CIE Reports

No later than July 21, 2008, the CIE shall provide via e-mail the final independent CIE reports and the CIE chair's summary report to the COTR William Michaels (William.Michaels@noaa.gov) at NOAA Fisheries. The COTR and alternate COTR Dr. Stephen K. Brown (Stephen.K.Brown@noaa.gov) will review the CIE reports to determine that the Term of Reference was met, notify the CIE program manager via e-mail regarding acceptance of the reports by July 28, 2008, and then distribute the reports to the NEFSC contact person.

Appendix 6 contd.

ANNEX 2 to Statement of Work: Contents of SARC CIE Independent Reports

1.

For each assessment reviewed, the report should address whether each Term of Reference of the SAW was completed successfully. For each Term of Reference, state why that Term of Reference was or was not completed successfully. To make this determination, CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable.

The report may include recommendations on how to improve future assessments.

The Independent CIE Report might also be used to provide greater detail than the SARC Summary Report on specific Terms of Reference or on additional questions raised during the meeting.

2.

If any existing Biological Reference Point (BRPs) proxies are considered inappropriate, include recommendations and justification for alternative proxies. If such alternatives cannot be identified, then indicate that the existing BRPs are the best available at this time.

3.

Any independent analyses conducted by the CIE reviewers as part of their responsibilities under this agreement should be incorporated into their Independent CIE Reports. It would also be helpful if the details of those analyses (e.g, computer programs, spreadsheets etc.) were made available to the respective assessment scientists.

4.

Additional questions that were not in the Terms of Reference but that are directly related to the assessments. This section should only be included if additional questions were raised during the SARC meeting.

5. The report shall include a list of all background material provided, a copy of the Statement of Work with Terms of Reference, and meeting agenda attached as separate appendices.

Appendix 6 contd.

ANNEX 3 to Statement of Work: Contents of SARC Summary Report

1.

The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether each Term of Reference of the SAW was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

To make this determination, the SARC chair and CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If the CIE reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

The report may include recommendations on how to improve future assessments.

2.

If any existing Biological Reference Point (BRP) proxies are considered inappropriate, include recommendations and justification for alternative proxies. If such alternatives cannot be identified, then indicate that the existing BRPs are the best available at this time.

3.

The report shall also include the bibliography of all materials provided during SAW 47, and any papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

The report shall also include as a separate appendix the Terms of Reference used for SAW 47, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.